THE RELATIVE INTELLIGIBILITY OF SPEECH RECORDED SIMULTANEOUSLY AT THE EAR AND MOUTH

DISSERTATION

Presented in Partial Fulfillment of the Requirements
for the Degree Doctor of Philosophy in the
Graduate School of The Ohio State
University

By

HERBERT JOSEPH OYER, A.B., M.S. in Ed.

The Ohio State University
1955

Approved by:

Henry M. Moser
Adviser
Department of Speech

Acknowledgments

The writer wishes to express his gratitude to Dr. Henry M. Moser, Adviser, and to Dr. John J. Dreher, Chief Investigator R F Project 519, for their guidance, encouragement and use of equipment during the course of this experiment.

The writer also wishes to acknowledge Dr. Dorothy Sherman for her guidance in the analysis and interpretation of the data.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D.)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

RELEVANT BACKGROUND MATERIAL

In communication involving speaking and listening, the usual mode of transmission is by air from the speaker's mouth to the listener's ear either with or without electronic amplification.

Some attempts have been made to determine how effectively vocalization can be picked up by stethoscopes and microphones from various areas of the body.

West,[1] in an early study stated that when a stethoscope is laid over any part of the known resonance area of the speech apparatus, larynx, throat, cheeks, floor of the mouth, nose, etc., a distinct vibration is communicated to the ear of the observer when the underlying resonating chamber is functioning.

Simon and Keller[2] describe in detail how they care-

[1]Robert West, "The Nature of Vowel Sounds," Quarterly Journal of Speech Education, 12, 1926, 244-293.

[2]Clarence Simon and Franklin Keller, "An Approach To The Problem of Chest Resonance," Quarterly Journal of Speech Education, 13, 1927, 432-439.

fully examined seven areas of the body and photographically recorded the vibrations as the subject intoned the vowel (o) at 128 dv. Voice and chest waves were recorded simultaneously by employing two phoneloscopes. They removed the diaphragm from a carbon microphone and made contact with the body by means of the pin which protruded one millimeter from the circular frame of the transmitter. By this procedure, vibrations were picked up within a restricted area of three inches around the point of contact. Analysis revealed that the frequency of vibration of the bony parts tended to agree with the cord tone, and that the greater the distance from the vocal cords the greater the variation in frequency of bone vibration. In all subjects, the right wing of the thyroid cartilage followed most closely the cord tone whereas the seventh rib followed the least. The frequency of vibration of the bony areas was that of the fundamental of the voiced tone.

Wise,[3] in an investigation of chest resonance shows

[3]C. M. Wise, "Chest Resonance," The Quarterly Journal of Speech, 18, 1932, 446-452.

the decrease in conductile efficiency among different types of tissue, in the following sequence: (1) bony tissue, (2) tendinous tissue, (3) tense muscle tissue, (4) relaxed muscle tissue, (5) soft non-muscular tissue.

Mullendore,[4] in studying the relative amplitudes of

[4]James M. Mullendore, "An Experimental Study of The Vibration of the Bones of the Head and Chest During Sustained Vowel Sounds," Speech Monographs, 16, 1949, 163-177.

sound vibrations at various body locations, shows the composite rank of intensity at ten microphone positions to be in the following order: (1) thyroid cartilage, (2) mandible, (3) nose, (4) top of head, (5) clavicle, (6) vertebra, (7) sternum (superior end), (8) sternum (inferior end), (9) mastoid, (10) rib (fifth).

Bekesy and Rosenblith[5] determined that the amplitude

[5]Georg von Békésy, Walter A. Rosenblith, "The Mechanical Properties of The Ear," Handbook of Experimental Psychology, Chapter 27, Ed. S.S.Stevens, (John Wiley and Sons Inc.: New York, 1951) pp. 1075-1115.

of the skin surface vibrations in the vicinity of the ear was only one-twentieth of the amplitude recorded near the vocal cords. They state that the vibrations of air in the oral cavity are transmitted from the cheeks to the lower jaw and that the attenuation between the oral cavity and the ear canal was forty to fifty decibels.

As indicated by the studies cited above, the literature provides examples of attempts made to record and analyze vibrations occurring at various body locations. The writer, however, was unable to find evidence in the literature of attempts made to record speech at the ear. Hence a canvass was made of possible sources of information regarding research directed toward recording speech at the ear. As a result of the canvass, research by Hirsh and Benson,[6] Schubert,[7] and Dolch[8] was brought to the attention

---

[6] I. J. Hirsh, R. W. Benson, "Acoustic Properties of Certain Oxygen Equipment, WADC Technical Report 52-175, Wright Air Development Center, May,1952, Appendix B, pages 20, 21.

[7] Earl D. Schubert, "Research Study of The Psycho-Acoustic Effects of Human And Artificial Sidetone," Final Report, Project No. 17-132 B. University of Iowa, Iowa City, Iowa, Pages 1-13.

[8] John P. Dolch, Earl D. Schubert, "Study of Body-Conducted Sidetone," Supplementary Report No. 6, Project No. 17-132 B., October, 1954. University of Iowa, Iowa City, Iowa. Pages 43-50.

---

of the writer.

Some preliminary experimentation involving analysis of the spectrum composition of speech energy in the ear canal has been carried out by Dr. Henning von Gierke at the Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio.

Hirsh and Benson state that a speech signal picked up from the ear canal was observed in their laboratories while attempting to test the A-13-A microphone, and that the ear canal does provide a source of sound pressure which can be utilized for the delivery of speech from talker to listener in a communications system. They assert that sounds from this source appear to have as good quality as the sounds picked up at the lips of the talker. This system, they state, is effective at fairly high noise levels and it is possible that required intelligibility can be obtained in ambient noise fields up to one hundred twenty decibels. They state in the report that several possibilities exist as to the actual source of sounds picked up at the ear. The first possibility is that of sound transmission around the head and through the ear cushion. A second possibility is via the bones of the head. The third possibility is the generation of sound pressure in the ear canal serving to activate the earphone and thus cause it to function as a microphone. They add that,

> Previous experiments done at this laboratory in June, 1950, show conclusively that the sound pressure is generated in the ear canal when the head is excited with bone conduction or by speech. In order to substantiate this theory, a test was performed by occluding the ears of the talker with ear plugs. This experiment showed a marked reduction in the intensity of sound.

Schubert[9] states that some investigators have observed

---

[9]Schubert, "Research Study of the Psycho-Acoustic Effects of Human and Artificial Sidetone," page 5.

an intelligible signal which can be picked up seemingly from the ear of the speaker. He adds that it is not difficult to demonstrate increase in sound pressure picked up by a probe placed under a conventional earphone when the speaker vocalizes while wearing the earphone. He adds still further that it is equally easy to demonstrate that inserting an ear warden into the external auditory meatus does not decrease the sound pressure picked up by the probe tube. He concludes that the component cannot be traced to any signal "coming out of the ear."

The writer has made numerous attempts to determine whether attenuation of sound occurs by occlusion of the external auditory canal with a plug (Ear Warden Type V-51-R). Conflicting evidence was obtained which could be used to support findings of either Hirsh or Schubert.

Dolch and Schubert[10] state that the source of the

---

[10] Dolch and Schubert, "Study of Body-Conducted Side-tone," pages 43-50.

---

sound pressure in the ear is still unknown, however they suggest that the contributing sources might be a function of (1) speech sound energy leaking through or around the ear-cushions, (2) sound pressure that is actually "emitted from the ear." These two possibilities are not held in serious regard by them inasmuch as attenuation of sound at the mouth does not decrease the level in the cavity. Also when a second person produces vowel sounds at a distance from the probe under the ear cushion comparable to the distance from the original speaker's mouth to ear, the frequency spectrum and sound pressure level is greatly altered. Failure to obtain decrease in sound pressure level by insertion of a plug into the ear canal causes them to reject "ear emission."

A third suggested contributing source was the possibility that a probe tube coupled to a microphone in a closed cavity produced an artifact. Probe-cavity interaction producing a spurious boost in low frequencies was discounted because previous measurements made in cavities of comparable size did not reveal this to be true. In addition, the probe-cavity interaction was not considered a valid contributing source since the same effect occurred when the probe was not employed.

The two Iowa investigators assert that a fourth possibility is one which merits serious consideration, viz. that the skull covered by the ear cushion is the vibratory source. This possibility is strongly reinforced by measurements made on other portions of the head and neck under earphone with probe which also reveal pressure variations of relatively similar magnitudes. The researchers conclude that the sound pressure source observed in the ear canal "is the vibrating enclosed portion of the head."

Regardless of the nature of the signal at the ear, the fact remains that an intelligible signal can be picked up there. This phenomenon can be demonstrated not only electronically but also mechanically.

In order to demonstrate that an intelligible speech signal could be mechanically transmitted from ear of talker to ear of listener, a crude instrument was constructed. The device consisted of two sets of stethoscope binaurals connected by five feet of glass and rubber tubing having a one-fourth inch bore. The experimenter placed the ear pieces of one set of binaurals in his ears in such a position that the connecting tube led from the back of his head to the ears of the listener. To prevent the oral signal of the speaker from being heard by the listener, the speaker placed himself in a sound treated room and attenuated the signal at the mouth with a sound baffle. The baffle consisted of a cardboard box (11" X 12" X 18") lined and filled with felt sound proofing material. The glass and rubber tubing was led through the wall of the sound treated room into an adjoining room. It was not difficult to demonstrate that the attempt to transmit intelligible speech from the ear of the talker to the ear of the listener was possible without electronic amplification.

As a result of publicity given to the above demonstration, George O. Tapper (formerly Director of the Signal Corps Acoustical Laboratory, Ft. Monmouth, N.J.) directed attention to earlier investigations concerning research related to the present problem. In the period of 1932 to 1934 the Signal Corps Laboratory was interested in developing a microphone that could be used in noisy environments and could be worn in a manner to leave both hands free. The late George A. Graham, Section Chief at that time, suggested that a sound powered receiver unit might be employed as a receiver and as a microphone driven by the sound that was transmitted through the Eustachian tube. All of the files covering this research have been retired but information has been obtained from communications of men who were associated with the research.

Harry W. Parmer, Chief, Advanced Systems Engineering Group, recalls that a bone conduction receiver unit was modified to provide a small diaphragm for driving the receiver unit. This diaphragm, when properly placed in close contact with the skin was fairly effective in excluding external noise. The highest output was obtained when the unit was used as a throat microphone, however, the best speech quality was derived when the unit was located either on the cheek close to the top of the ear or on the cheek directly in front of the lower part of the ear. Further consideration of the cheek units was abandoned due to mounting difficulties.

Excellent speech signal quality was obtained by the Signal Corps investigators when using a sound powered unit connected in a push-to-talk circuit with an appropriate amplifier. The greatest limitation was the relatively low level output from the sound powered unit which was used at the ear. The amplifier was required to have a gain of approximately 80 db, which meant the sound level in the ear was down approximately 60 db from that obtained in front of the mouth.

Albert E. Woodruff (then a member of the Research Department, Automatic Electric Company, which was requested by the Signal Corps to design an especially sensitive receiver for use as an ear microphone) recalls the early demonstrations by Graham and the work of testing the special receiver. He noticed that if one did not press the receiver tightly against the ear the output was increased. This was taken as evidence that the receiver was getting its stimulus not through the ear mechanism but from the speech organs through the air. Better ear cushions and other means of shutting out air-borne sound confirmed this. He concludes that no more speech sound could be picked up in the ear canal than could be picked up from the vibrations of the skull elements surrounding it. He states that there is no useful transmission of speech through the Eustachian tubes or other soft membranes of the head. When given 30 to 50 db of amplification, speech can be picked up by a proper transducer in contact with any part of the skull, the throat or upper thorax.

The work of the Signal Corps project resulted in the adoption of the throat microphone.

THE PROBLEM

The problem under investigation in this study was to determine the relative intelligibility of simultaneously recorded speech signals picked up at the lips and left ears of speakers. The samples of speech picked up at both positions were reproduced for a group of listeners in varying levels of white noise. Listeners wrote down the words they heard. The relative intelligibility of speech picked up at both positions was compared.

The following questions were proposed: (1) Are speech signals, origin ear recorded in quiet as intelligible as speech signals, origin lips at each of the three signal-to-noise ratios employed in the test?, (2) Is the trend for intelligibility over the S/N ratios employed in the test in the same direction for speech signals of both origins?,

(3) Are speech signals of ear origin recorded in quiet as intelligible in quiet as speech signals of mouth origin?

IMPORTANCE OF THE STUDY

Pickup of the speech signal at the ear might prove to be advantageous for a number of reasons: (1) in aircraft communication the microphone could be separated from the oxygen system, thus eliminating pickup of oxygen valve flutter, an interfering low frequency noise, (2) reaching for the swing boom supported microphone could be eliminated when flying at low altitudes not requiring the use of the oxygen system, (3) breath noises as a result of pressure breathing could be lessened. Hirsh[11] states,

---

[11] Hirsh, _op_. _cit_., page 21.

> The use of this type of system would separate the communications system from the oxygen system and would therefore have an advantage over the present location of the microphone.

CHAPTER II

PROCEDURE

The procedures followed in this investigation contain certain refinements brought about as a result of a preliminary study of similar magnitude. Reference to the earlier experimentation is made later in this chapter.

Selection of material. The fifty monosyllabic words employed in this experiment were selected from a list of words used in air traffic control. This list was compiled in 1951 by M. Y. McCormack (Chief Psychologist, Civil Aeronautics Administration) from information obtained from one hundred twenty-five C.A.A. control towers. Twelve of the towers were international, *i.e.* handling incoming foreign airline traffic. Each installation contributed five pages of communication from tower to plane. This resulted in a tabulation of one hundred twenty-one thousand two hundred ninety-six words. Twenty-five per cent of the words were numbers, the remaining seventy-five per cent were operational terms and geographical names.

The fifty-word list employed in the experiment was selected to represent vowel, diphthong, and initial consonant sounds in relative proportion to the frequency of their occurrence in English monosyllabic words. The word list, with frequency of use in air traffic is presented in Appendix B.

Selection of speakers. Six adult males served as speakers. One was a Professor of Speech. Five were graduate students in Speech Science. All were of General American Dialect.

Instruction to speakers. Speakers were instructed to read five randomizations of the fifty-word list in the following manner:

Number one you will write ________ ________ ________
Number two you will write ________ ________ ________
etc.

Speakers were asked to pause approximately one second between the last word of the carrier phrase and the first stimulus word, and two seconds between ensuing stimuli of the unit. They were instructed to maintain level inflection on all stimulus words.

Microphones. Inasmuch as there were no available microphones built specifically for the ear, the experimenter employed two Dyna-Lab D-69 Magnetic Insert Earphones, one at the ear, and the other at the lips of speakers. Choice of this particular transducer was made after surveying response curves of a number of commercial models.

Recorders. Two tape recorders (Ampex, Models 600, 305) were employed.

Recording procedure. One transducer was mounted on a standard and positioned three centimeters from the lips of the speaker. The other transducer was coupled to a custom fitted ear mold in the left ear of the speaker. The ear signal was fed to the Ampex, Model 600 and the mouth signal to the Ampex, Model 305. In this manner, signals picked up at the ear and mouth were simultaneously recorded. Tape speed was seven and one-half inches per second for each recorder.

Preparation of the training tape. Twelve word lists were randomly transcribed to a training tape. The training tape contained one hundred stimulus words from each of the six speakers (fifty ear and fifty mouth) making a total of six hundred stimuli for training purposes. Tape speed was seven and one-half inches per second.

Preparation of the experimental tape. Thirty-six speaker lists were randomly transcribed to a master tape. Thus for one listening condition, one ear and one mouth list from each of the six speakers was represented. One restriction was placed on the randomization. No two lists (ear and mouth) from the same speaker appeared adjacent to each other. In making the transcription, care was exercised in setting the level of the last word of each carrier phrase between -1 and -2 VU.

The listeners. Twenty-four American-born university students served as listeners. All listeners were given an audiometric sweep check of hearing. A D-9 Maico audiometer was employed. The hearing acuity criterion set was + 15 db or better from 250 through 8000 cps.

Training the listeners. Since there were no known intelligibility values established for the stimulus list, listeners were trained to recognize the words. The experimenter read aloud the fifty-word list in free field as the listeners followed a printed list. The training tape was played back on an Ampex Tape Recorder (Model-305) and fed through an electronic attenuation network to earphones (Permaflux, type PDR-8).[12] Facilities in the laboratory permitted only twelve listeners to participate at a time. Each listener was given one and one-half hours of training in quiet and three hours in varying levels of noise.

Administering the test. The test tapes were played back through the same system as the training tape. White noise, produced by a Harvard type, magnetically stabilized gas tube generator was mixed with the signal at the output of the attenuator. The white noise was kept constant at a level of approximately 74 db. Since the amount of the

---

[12]Schematic of playback equipment is presented in Figure 1, page 17.

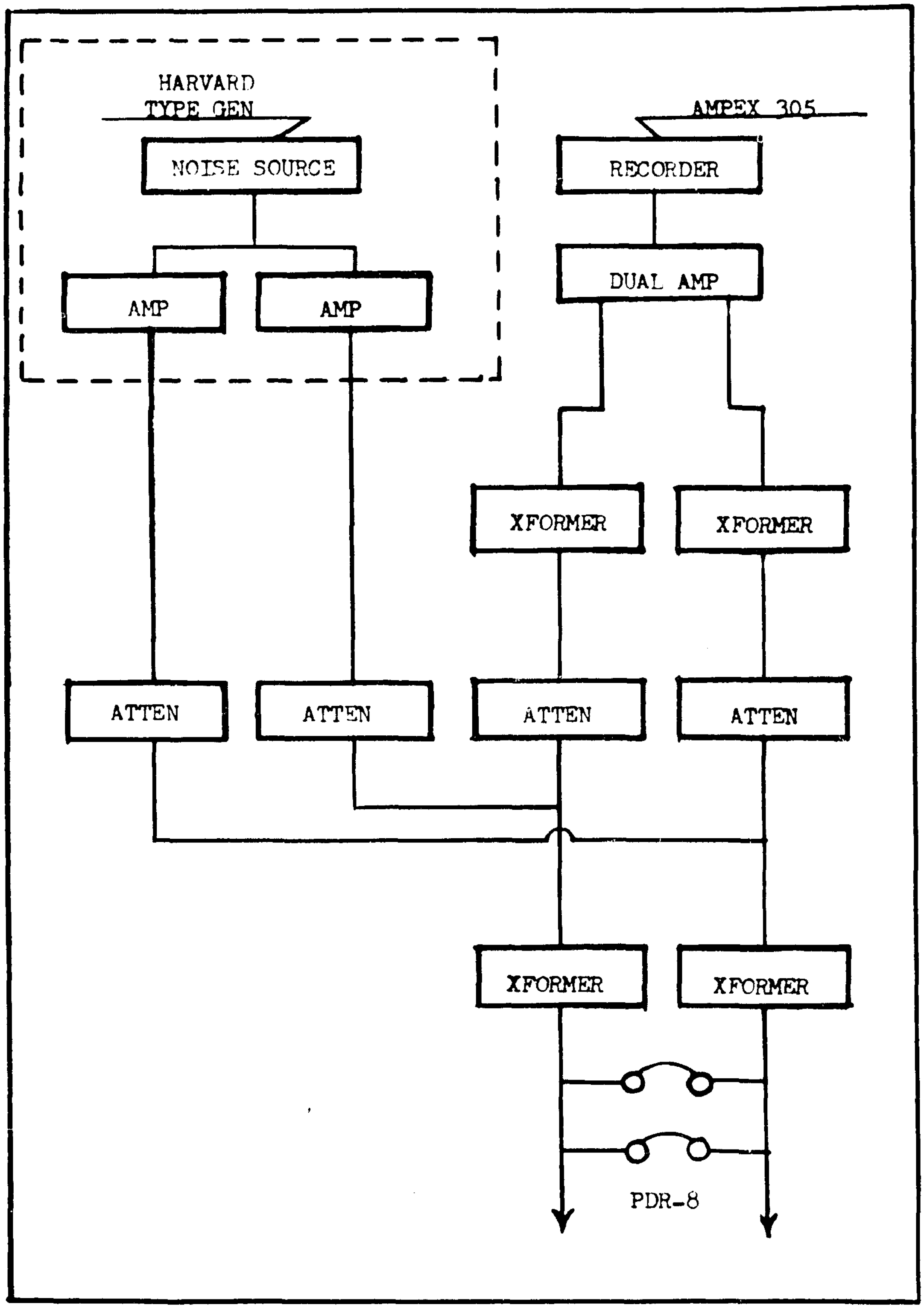

Figure 1

Block Diagram of Playback Equipment

attenuation could be varied in steps of one decibel, the desired signal-to-noise ratio could be obtained. Three S/N ratios were employed as test conditions (-12, -15, and -18 db S/N).

As a part of each test session, listeners warmed up to the task by writing down a thirty-six word list made up of stimulus words from each speaker. Ear and mouth signals were equally represented.

A thirty-five minute test session was followed by a fifteen minute rest period. A total of seventy minutes of listening was required of each listener per day. Total test time for each subject was three and one-half hours.

Rationale underlying certain aspects of procedure. The rationale underlying certain aspects of procedure employed in this investigation grew out of preliminary experimentation. In the preliminary investigation, the transducer was imbedded in a clay matrix and suspended at the entrance of the ear canal with only the lower rim of the circular frame of the transducer in contact with the superior surface of the anti-tragus. The transducer and clay were covered by a webbed flight helmet and a modified crash helmet worn by the speaker. Intelligibility testing of signals recorded by this arrangement revealed, that, although fairly intelligible in quiet, the ear signal did not compare favorably with the mouth signal when tested under conditions of noise. Therefore a critical evaluation was made of the procedure with reference to "microphone" placement at the ear. The idea was that a more intelligible signal could be realized with a more advantageous placement. Therefore the transducer was coupled to a custom fitted ear mold. Preliminary tests with a fitted ear mold indicated that the use of the clay seal and helmets was not only impracticable but perhaps undesirable as well. A series of tests employing the best available means for isolating the voiced signal indicated that apparently enough energy was being transmitted via bone to make isolation of the ear signal unnecessary. It was further noticed that the intelligibilities of signals simultaneously recorded at mouth and ear under this system compared very favorably. Breath noises picked up by the mouth "microphone" were completely eliminated in the pickup of the bone transmitted signal at the ear.

The longer training period given to listeners stemmed from inspection of intelligibility scores made in the preliminary experiment. Even though the ninety-six per cent intelligibility criterion was met in quiet in the preliminary study, it appeared that learning was taking place when the test signals were presented in noise. Hence the training period was increased from two hours in quiet to four and one-half hours (one and one-half in quiet and three hours in varying levels of noise).

CHAPTER III

RESULTS AND DISCUSSION

A triple analysis of variance was employed to evaluate the data. Factors analyzed were S/N ratios (R), origins (O), and listeners (L). The S/N ratios were (1) -12 db, (2) -15 db, and -18 db. Origins of speech signal pickup were (1) mouth and (2) ear.

The criterion measure was the total number of correct responses to the six speakers for one listener at one S/N ratio. A correct response was one in which the stimulus word or a homophone was reproduced in written form. The table of criterion measures is presented in Appendix B.

Results of the triple analysis of variance are presented in Table I. The <u>F</u>-test of interaction between S/N ratios and origins is significant beyond the 1 per cent point. Tests of simple effects were made for (1) origin mouth, (2) origin ear, (3) -12 db S/N ratio, (4) -15 db S/N ratio, and (5) -18 db S/N ratio. Results of the analysis of simple effects are presented in Table II. All simple effects were significant beyond the .01 point except the difference between origins at -12 db S/N ratio, which was significant at the 6 per cent point.

Mean values for each origin and each ratio are presented in Table III. Mean square values from Table I were

Table I

Results of Analysis of Variance For Main Effects

Source	df	ss	ms	F	P
Ratios	2	347003.60	173501.80		
Origin	1	36544.69	36544.69		
Listeners	23	42346.22	1841.14		
R X O	2	21930.10	10965.05	118.40*	.01
R S L	46	9362.74	203.54		
O X L	23	1860.98	80.91		
R X O X L	46	4260.23	92.61		
Total	143				

$*\underline{F} = \frac{ms_{RxO}}{ms_{RxOxL}}$; $\underline{F}_{.01} = 5.10$

Table II

Summary of Analysis of Simple Effects of S/N Ratios and Origins

Source	df	ss	ms	F^1	P
S/N Ratio -12	1	379.69	379.69	4.69	.06
S/N Ratio -15	1	7525.02	7525.02	93.00	.01
S/N Ratio -18	1	50570.09	50570.09	625.00	.01
$\text{Origin}_{\text{mouth}}$	2	270741.00	135370.50	665.08	.01
$\text{Origin}_{\text{ear}}$	2	98192.69	49096.35	241.25	.01

[1]Error terms for computing $\underline{F}$ are from the analysis summarized in Table I:

$\underline{F}_{R(-12,-15,\text{ or }-18)} = ss_{R(-12,-15,\text{ or }-18)} / ms_{RxL}$, df = 1 &46

$\underline{F}_{O(-12,-15,\text{ or }-18)} = ss_{R(-12,-15,\text{ or }-18)} / ms_{RxL}$, df = 2 &23

employed to compute differences required for significance.[13]

---

[13]Required difference = $\underline{t}\ (2\ ms_{error}/n)^{1/2}$

Examination of mean values together with differences required for significance indicates that (1) differences in intelligibility in favor of the ear exist at the three S/N ratios employed, (2) at -15 and -18 db S/N ratios, the ear signal is significantly better recognized, (3) speech picked up at mouth and ears of speakers becomes significantly less intelligible at each S/N ratio, and (4) the decrease in the mean values for origin ear through increasing S/N ratios is less than those for origin mouth.

Table III

Mean Values for Origins for Each S/N Ratio

Origin[1]	S/N Ratios[2] -12 db	-15 db	-18 db
Mouth	199.88	135.13	50.13
Ear	205.50	160.17	150.04

[1]As indicated by F-tests of Table II the difference between origins was significant at the 6 per cent level for -12 db S/N ratio and beyond the 1 per cent level for -15 and -18 db S/N ratios.

[2]c.d. = $\underline{t}_{.01}\ (2ms_{RxL}/n)^{1/2} = 10.60$

DISCUSSION

In view of the questions proposed at the outset of the experiment, the analysis of data points out that within the experimental confines of this study (1) speech signals, origin ear recorded in quiet, are more intelligible than simultaneously recorded speech signals, origin mouth, at each of the three S/N ratios employed in the test, (2) the trend for intelligibility over the S/N ratios employed in the test is in the same direction for signals of both origins. Although the trend is in the same direction, the intelligibility scores for the mouth signals decrease much more rapidly than intelligibility scores for ear signals. Inspection of Table III reveals that ear signals when heard at -18 db S/N ratio are significantly better recognized than mouth signals when heard at -15 db S/N ratio.

Although a condition of quiet was not included as a test condition for statistical analysis, it can be stated that speech signals of ear origin are as intelligible in quiet as speech signals of mouth origin by virtue of the fact that the number of correct responses to ear signals was only one less than for mouth signals when the training tape was played in quiet. The training tape was played only once in quiet and contained an equal number of stimuli from both origins with all speakers represented, as did the test tapes.

It will be recalled that the speech signal picked up at the ear was by means of a transducer coupled to a custom fitted ear mold. Hence the entire surface of the mold, in contact with the skull, served as a transmitter of bone vibrations which occurred as the speaker phonated. With this arrangement, full advantage was taken of bone transmitted vibrations in the ear canal. It will also be recalled that in preliminary testing an intelligible ear signal of good quality was observed when the signal at the mouth was isolated and the clay seal was removed from the ear. As a result of this observation the ear seal was considered impracticable. Inasmuch as the seal was not employed, it was recognized that the plastic case surrounding the transducer and the mold itself were capable of transmitting air-borne sound. It seems reasonable to assume then, that the superiority of the ear signals was in part a function of combined air and bone vibrations. The bone vibration component however was the major contributing source as shown by preliminary laboratory tests in which an intelligible signal of good quality was delivered from a sealed ear in an ambient noise field of 90 db while the mouth signal was attenuated by a baffle box. It is important to add at this point that the removal of the ear seal and the baffle at the mouth while recording in ambient noise, results in an almost complete masking of the ear signal. However when the mouth baffle is held in position and the ear seal removed, the ear signal recording in noise is highly intelligible. Laboratory observation of this phenomenon underlies the writer's reasoning, viz. that the major source of energy in the air-bone coupling is the bone component.

Dolch and Schubert[14] suggest two possible explanations

[14]Dolch and Schubert, "Study of Body-Conducted Side-tone," page 29.

for increase in overall sound pressure at the ear when the mouth box is in position. One possibility is that normal patterns of skull and jaw vibration are substantially altered by tight coupling of the mouth box to the face. The other possibility is that the ratio of radiated to reflected energy is altered due to the change in acoustic impedance presented to the vocal tract as a result of the box coupled to the mouth.

Upon listening to recordings of mouth signals it was noticed that breath noises were present in the recordings of some speakers. Interfering breath noises were not apparent in recordings of ear signals. The superiority of the ear signals was in part apparently a function of the absence of interfering breath noises.

The results of this study indicate that the ear signal is worthy of serious consideration in aircraft communication.

However, it appears at this point that practical utilization of an ear signal recorded in noise is contingent upon the development of a suitable noise shield or a noise cancelling ear microphone. Preliminary work involving comparisons of shields of various materials and enclosing various volumes seems to indicate very satisfactory performance when the transducer is protected by a thin plastic cup which does not bear upon either the pinna or the transducer. The most satisfactory seal tested had a volume of 220 cc. No systematic attempt at determining an optimum volume was made although it was felt that the size of the cavity could be a highly important factor, at least with the transducer tested.

CHAPTER IV

SUMMARY AND CONCLUSIONS

SUMMARY

The purpose of the study was to determine the relative intelligibility of simultaneously recorded speech picked up at the lips and left ears of six adult male speakers. Test material consisted of fifty monosyllabic words selected from words used in air traffic control. Vowel, diphthong, and initial consonant occurrence in the test list approximated the relative frequency of their occurrence in English monosyllables. Each speaker recorded three randomized lists. These lists were randomly presented (with respect to origin and speakers) to a panel of twenty-four trained listeners at -12, -15, and -18 db S/N ratios. The criterion measure employed was the total number of correct responses to six speakers for one listener at one S/N ratio.

Data obtained were analyzed by the analysis of variance technique. Origin of speech signal, S/N ratios, and listeners constituted the three factors in the design.

CONCLUSIONS

Analysis of results obtained within the framework of the investigation indicate the following conclusions:

1. When heard in quiet, speech picked up by a transducer coupled to a custom fitted ear mold (in the ear of the speaker) is as intelligible as speech recorded simultaneously at the lips of the speaker by a second transducer (same model as at the ear).
2. Speech signals from the ear are more intelligible than speech signals from the mouth for all S/N ratios employed, with increasingly large differences at lower ratios.
3. The trend for intelligibility over S/N ratios employed in the test is in the same direction for speech signals of both origins. Decreasing S/N ratio is significantly destructive to the intelligibility of speech picked up at the ear and lips of speakers, and is more destructive to the latter.

APPENDIX A

FIFTY MONOSYLLABIC WORDS

Alphabetical listing	Frequency	listing
BASE	ONE	4955
BEACH	THREE	4090
BLOCK	FIVE	3080
BRIDGE	EIGHT	2077
CALL	NINE	2242
CHANGE	FOUR	2058
CLEAR	SIX	2014
CRUISE	THIS	1117
DASH	TAKE	1080
DEW	LAND	1015
DOWN	RIGHT	952
EAST	TURN	908
EIGHT	SOUTH	565
FIVE	STRIP	193
FOG	UP	185
FOUR	RUN	122
FROM	WHAT	120
FUEL	POINT	114
GOOD	LEG	98
GUARD	HOLD	98
HALF	GUARD	86
HOLD	FROM	78
JET	LEFT	77
LAND	CRUISE	76
LEFT	CLEAR	65
LEG	DOWN	60
MILE	STAND	60
NINE	BASE	59
ONE	EAST	58
PASS	BEACH	41
POINT	CALL	40
RAIN	FOG	35
RIGHT	HALF	30
RUN	CHANGE	28
SHORT	SHORT	28
SIX	DEW	23
SNOW	MILE	23
SOUTH	SNOW	22
SMOKE	PASS	21
STAND	RAIN	20
STRIP	SMOKE	18
TAKE	DASH	10
THIS	BRIDGE	7
THREE	GOOD	6
TURN	FUEL	6
UP	JET	4
WHAT	BLOCK	4
WING	YES	4
YES	WING	3
ZONE	ZONE	2

APPENDIX B

Criterion measures at -12,
-15, and -18 db S/N
ratios

-12 db S/N Ratio

Listen-ers	Speakers 1		2		3		4		5		6	
	M	E	M	E	M	E	M	E	M	E	M	E
1	43	41	39	29	30	39	41	42	22	37	43	43
2	37	36	37	28	27	41	37	31	20	34	39	48
3	42	35	38	33	29	29	39	34	27	28	38	46
4	43	41	36	29	33	46	40	35	23	39	43	47
5	40	33	38	30	28	39	36	38	18	32	37	47
6	41	35	41	39	32	40	41	37	23	35	39	46
7	39	38	39	16	25	28	27	25	23	24	41	41
8	40	39	36	28	32	38	36	33	21	31	41	45
9	37	36	25	20	17	30	28	28	18	26	38	43
10	35	33	34	25	28	38	32	33	20	32	35	43
11	42	37	38	30	28	35	35	36	25	31	36	46
12	43	35	38	21	28	30	31	31	26	30	32	40
13	34	40	35	24	23	29	37	34	25	25	36	45
14	41	32	36	23	34	37	35	30	24	32	32	46
15	39	31	36	28	27	33	42	32	17	30	32	43
16	45	35	38	28	31	34	34	33	25	35	41	43
17	34	28	34	26	28	34	33	28	24	25	33	45
18	39	32	33	26	28	35	38	28	24	30	39	42
19	36	37	38	29	30	33	32	29	23	35	36	40
20	38	38	37	25	31	36	42	30	22	37	35	42
21	43	29	39	24	31	39	33	33	20	31	36	46
22	43	39	37	30	35	37	33	36	25	34	36	46
23	39	35	33	19	19	32	33	33	20	25	36	43
24	41	34	39	30	31	35	35	39	27	28	38	46

-15 db S/N Ratio

Listeners	Speakers 1 M	1 E	2 M	2 E	3 M	3 E	4 M	4 E	5 M	5 E	6 M	6 E
1	27	33	37	36	15	26	24	31	24	36	36	43
2	32	26	35	24	15	24	25	30	14	28	30	38
3	25	20	32	27	13	19	27	28	15	25	24	41
4	22	37	39	35	18	26	30	31	17	35	32	39
5	25	36	36	22	12	24	31	33	14	38	24	44
6	23	27	27	23	21	27	30	28	21	31	30	31
7	15	22	30	19	10	24	14	23	13	19	25	34
8	18	26	33	26	14	28	30	18	14	30	23	38
9	11	20	29	14	5	11	26	28	12	24	19	33
10	22	22	30	24	15	20	28	25	12	26	24	35
11	23	31	42	25	13	25	27	30	20	28	26	39
12	15	24	30	20	8	24	23	21	12	29	24	31
13	24	20	37	19	13	20	32	27	19	24	30	25
14	24	26	25	18	12	19	24	27	17	28	34	33
15	21	21	37	20	16	18	29	29	22	21	26	33
16	27	33	33	35	16	24	31	32	23	31	39	39
17	13	27	30	22	6	23	16	27	14	25	20	32
18	24	27	37	21	13	24	25	30	13	30	30	37
19	26	25	34	23	19	18	25	28	9	21	29	36
20	22	23	32	22	10	17	23	25	16	22	31	35
21	17	21	33	17	10	17	19	27	13	24	32	32
22	18	22	33	28	11	24	20	24	15	32	32	35
23	15	18	24	14	9	19	24	31	13	18	22	30
24	19	22	30	21	14	24	19	25	14	25	23	34

<u>-18 db S/N Ratio</u>

Listen-ers	Speakers 1		2		3		4		5		6	
	M	E	M	E	M	E	M	E	M	E	M	E
1	24	20	19	25	11	32	24	36	9	19	13	30
2	15	19	13	17	6	20	26	26	7	16	6	30
3	3	23	7	19	7	15	2	22	1	13	5	19
4	18	22	20	19	7	32	27	29	13	17	13	38
5	7	20	13	21	5	28	10	24	2	7	9	22
6	13	23	19	17	5	25	23	25	6	14	3	30
7	3	14	3	14	1	15	2	15	2	3	2	16
8	16	21	12	14	5	20	16	23	5	10	7	32
9	5	14	8	17	2	13	5	21	1	8	4	19
10	2	13	6	10	0	12	1	25	1	11	4	13
11	4	20	7	19	8	23	8	30	1	8	1	20
12	3	13	5	11	3	16	7	29	4	5	0	13
13	6	13	10	15	5	14	3	20	3	7	4	13
14	16	18	14	14	7	23	11	30	1	11	7	24
15	7	15	5	15	8	26	9	25	7	10	11	29
16	17	24	19	18	10	30	19	32	6	20	13	33
17	9	13	8	15	2	9	8	22	2	5	2	19
18	5	16	13	13	7	17	13	27	2	12	4	22
19	7	23	20	17	5	21	16	31	4	11	6	23
20	5	20	7	13	6	20	7	26	2	5	6	15
21	11	19	20	16	10	25	13	30	4	10	2	29
22	19	19	16	16	9	19	20	26	10	12	10	31
23	12	15	13	15	4	15	7	22	3	11	8	24
24	6	22	10	17	5	24	14	23	10	9	8	24

BIBLIOGRAPHY

SOURCES CITED

Békésy, Georg von, Walter A. Rosenblith, "The Mechanical Properties of the Ear," Chapter 27, Handbook of Experimental Psychology, Ed., S. S. Stevens, New York: John Wiley and Sons, Inc., 1951. 1436 pages.

Dolch, John P., Earl D. Schubert, "Study of Body-Conducted Sidetone," Supplementary Report No. 6, Project No. 17-132 B, Signal Corps, October, 1954. University of Iowa, Iowa City, Iowa. 60 pages.

Hirsh, I.J., R.W. Benson, "Acoustic Properties of Certain Oxygen Equipment," WADC Technical Report 52-175, Wright Air Development Center. May, 1952, 28 pages.

Mullendore, J.M., "An Experimental Study Of The Vibration Of The Bones Of The Head And Chest During Sustained Vowel Sounds," Speech Monographs, XVI (1949), 163-177.

Parmer, Harry W., Chief Advanced Systems Engineering Group, Fort Monmouth, New Jersey. Letter - February 21, 1955.

Schubert, Earl D., "Research Study Of The Psycho-Acoustic Effects Of Human And Artificial Sidetone," Final Report, Project No. 17-132 B, Signal Corps, University of Iowa, Iowa City, Iowa. 13 pages.

Simon, Clarence, Franklin Keller, "An Approach To The Problem of Chest Resonance," Quarterly Journal of Speech Education, XIII (1927), 432-439.

Tapper, George O., Formerly Director Acoustical Laboratory, Fort Monmouth, New Jersey. Letter - December 28, 1954.

von Gierke,Dr., Henning, Bio-Acoustics Section, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio. Letter - November 10, 1954.

West, Robert, "The Nature of Vowel Sounds," Quarterly Journal of Speech Education, XII (1926), 244-293.

Wise, C. M., "Chest Resonance," The Quarterly Journal of Speech, XVIII (1932), 446-452.

Woodruff, Dr. Albert E., 1033 West Van Buren St., Chicago, Illinois. Letter - January 20, 1955.

AUTOBIOGRAPHY

I, Herbert Joseph Oyer, was born in Groveland, Illinois, on January 28, 1921. I received my secondary education at Bluffton Richland High School, Bluffton, Ohio. My undergraduate training was obtained at The Ohio State University and Bluffton College. I received the degree Bachelor of Arts in 1943 from Bluffton College. In 1949 I received the degree Master of Science in Education from Bowling Green State University.

In 1946-47 I was employed as a high school teacher at Margaretta High School, Castalia, Ohio. While working toward the masters degree at Bowling Green State University in 1947-48, I served as a graduate assistant in the department of speech. From 1948-1953 I was a member of the faculty at Bowling Green State University.

Two years have been spent in residence at The Ohio State University specializing in the Speech Science area. The second year I have held the position of Research Assistant while completing the requirements for the degree Doctor of Philosophy.